GLUTAMINE:FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE (GFAT) COMPRISING AN INTERNAL PURIFICATION TAG, AND ITS USE FOR THE SCREENING OF COMPOUNDS

The present invention relates to a modified glutamine:fructose-6-phosphate amidotransferase, which can be rapidly purified in quantities sufficient for the screening of compounds modifying its activity.

Glutamine:fructose-6-phosphate amidotransferases (GFAT), EC 2.6.1.16, also called glucosamine-6-phosphate synthases or 2-deoxy-glucose-6-phosphate ketol isomerases, are involved in the biosynthesis route of hexosamines. GFAT catalyzes the first, limiting, stage of this biosynthesis route according to the reaction:

L-Glutamine + fructose-6-phosphate ⟶ L-Glutamate + glucosamine-6-phosphate

by transfer of the amidic nitrogen from the L-Glutamine to the ketone function of the fructose-6-phosphate. The GFATs therefore control the flow of glucose in the route of the hexosamines, via the fructose-6-phosphate, and consequently the formation of the hexosamines produced.

A recombinant bacterial form of GFAT, the glucosamine-6-phosphate synthase of *Escherichia coli*, has been purified to homogeneity and studied exhaustively. The properties and the enzymatic mechanism of the amide transfer have in particular been widely described (article by Teplyakov *et al.*, *Nat. Prod. Rep.* (2002) **19**:60). In particular, this enzyme, the crystalline structure of which has been resolved (Teplyakov *et al.*, *J. Mol. Biol.* (2001) **313**:1093), is formed by two domains, one having a hydrolase activity (glutaminase) and the other an isomerase activity.

Moreover, eukaryotic GFATs have been characterized, including in particular that of rat liver (Huynh *et al.*, *Arch. Biochem. Biophys.* (2000) **379**:307) and that of the yeast *Candida albicans* (Milewsky *et al.*, *J. Biol. Chem.* (1999) **274**:4000).

In humans, preliminary studies have shown the presence of GFAT activity in the liver (Ghosh *et al.*, *J. Biol. Chem.* (1960) **235**:1265). Several GFATs are now known. GFAT1, the principal form, GFAT2, which is preferentially expressed in the central nervous system, and GFAT1Alt, an isoform of GFAT1, essentially expressed in the striated muscles. The peptide sequences of GFAT1 and GFAT2 possess 75% sequence identity with each other, and those of GFAT1 and GFAT1Alt are identical except for an insertion of 18 amino acids into the GFAT1Alt sequence. The sequences of GFAT are therefore very preserved in humans, but also between species, since the peptide

sequences of human GFAT1 and *E. coli* GFAT or mouse GFAT1 have 35% and 99% identity respectively.

The human GFAT1 gene was cloned in 1992 (McKnight *et al.*, *J. Biol. Chem.* (1992) **267**:25208). It codes a protein of 77 kDa formed by two distinct domains (Teplyakov *et al.*, *Nat. Prod. Rep.* (2002) **19**:60).

The increase in the production of UDP-NAc-GlcNH$_2$, the final product of the biosynthesis route of the hexosamines, and its accumulation in the tissues have recently been involved in the development of insulin-resistance (Marshall *et al.*, *FASEB J.* (1991) **5**:3031, Yki-Jarvinen *et al.*, *Diabetes* (1996) **45**:302, Thompson *et al.*, *J. Biol. Chem.* (1997) **272**: 7759, Hawkins *et al.*, *J. Clin. Invest.* (1997) **99**:2173, Robinson *et al.*, *Diabetes* (1993) **42**:1333, Daniels *et al.*, *J. Clin. Invest.* (1995) **96**:1235, Baron *et al.*, *J. Clin. Invest.* (1995) 96:2792).

Thus, it has been shown that an increase in the cell level of UDP-NAc-GlcNH$_2$ by a slight overexpression of GFAT1, or a supply of exogenic glucosamine, can induce insulin-resistance both in vivo and in adipocytes in culture (Robinson *et al.*, *Diabetes* (1993) **42**:1333, Daniels *et al.*, *J. Clin. Invest.* (1995) **96**:1235, Baron *et al.*, *J. Clin. Invest.* (1995) **96**:2792, Hebert *et al.*, *J. Clin. Invest.* (1996) **98**:930).

In fact, insulin activates its transduction route by binding to its receptor, which induces the translocation of the glucose transporters, such as the GLUT4 receptor, stored in the cell, towards the membrane, and increases the inflow of glucose. The glucose thus enters the glycolysis route and is converted to glucose-6-phosphate then to fructose-6-phosphate. When the inflow of glucose is excessive, the fructose-6-phosphate enters the biosynthesis route of the hexosamines and is converted to glucosamine-6-phosphate by the GFAT. Several observations indicate that the metabolites of the glucosamine-6-phosphate prevent the translocation of the glucose receptors towards the cell membrane, thus reducing the inflow of the cell glucose (Marshall *et al.*, *FASEB J.* (1991) **5**:3031, Giacarri *et al.*, *Diabetologia* (1995) **38**:518, Marshall *et al.*, *J. Biol. Chem.* (1991) **266**:4706, Paterson *et al.*, *Endocrinology* (1995) **136**:2809).

The mechanism by which the metabolites of the glucosamine-6-phosphate exercise their physiological effects is not clear. One hypothesis has however been proposed: a high cytosolic concentration of UDP-NAc-GlcNH$_2$ would lead to the hyperglycosylation of the Ser or Thr phosphorylation sites, thus leading to the stopping of the insulin-signalling route (Comer *et al.*, *J. Biol. Chem.* (2000) **275**:29179).

The GFAT activity is therefore considered as being one of the causes of high levels of blood glucose; moreover it is known to be high in patients suffering from non-insulin-dependant sugar diabetes or type II diabetes (Yki-Jarvinen *et al.*, *Diabetes* (1996) **45**:302).

Obtaining GFAT inhibitors would make it possible to reduce glycaemia in particular in individuals suffering from pathologies linked to hyperglycaemia, such as type II diabetes, acidosis and/or diabetic ketosis, for example.

Fungal or plant GFAT inhibitors could also make it possible to obtain fungicides and herbicides respectively.

However, in spite of the obtaining of recombinant forms of GFAT, the instability of the enzymatic preparations obtained, their small quantity, and their insufficient purification level, have not made it possible to obtain effective GFAT inhibitors.

A subject of the invention is therefore to provide a modified GFAT the activity of which is stable and which can be obtained in a large quantity, with a high level of purity and of activity.

The present invention relates to an enzymatically-active protein comprising:

- a GFAT sequence and at least one purification tag sequence, the purification tag sequence being inserted between two consecutive amino acids of the GFAT sequence, or
- a sequence deriving from the preceding sequence by suppression, insertion or mutation of at least one amino acid, provided that said protein has an enzyme activity, or
- a sequence having at least 35%, in particular at least 90%, of sequence identity and/or at least 44%, in particular at least 95%, of sequence similarity with one of the preceding sequences, provided that said protein has an enzyme activity.

The term GFAT designates a class E.C. 2.6.1.16 enzyme catalyzing the following reaction:

L-Glutamine + fructose-6-phosphate $\longrightarrow$ L-Glutamate + glucosamine-6-phosphate

in particular under the experimental conditions as described in the example which follows or in Broschat *et al.*, *J. Biol. Chem.* (2002) **277**:14764.

GFAT is designated by the name of glutamine:fructose-6-phosphate amidotransferase, or also glucosamine-6-phosphate synthase or 2-deoxy-glucose-6-phosphate ketol isomerase.

The term "enzymatically-active protein" designates a protein having a catalytic action.

Advantageously, the enzymatically-active protein possesses a GFAT activity.

The term "purification tag" designates a peptide sequence capable of binding specifically to a given ligand. Advantageously, the binding of said ligand to the purification tag makes it possible to form a complex between the protein carrying the purification tag and said ligand, said complex being able to be specifically isolated.

Advantageously, the purification tags according to the invention are not placed at the end of the peptide chain, at the N-terminal or C-terminal end, but inside the peptide chain.

The term "sequence identity" designates the percentage of identical amino acids between two aligned sequences, in particular using algorithms such as that defined by Altschul *et al.*, *Nucleic Acids Res.* (1997) **25**:3389, for example.

The term "sequence similarity" designates the percentage of similar amino acids, i.e. amino acids the side chains of which possess similar physico-chemical properties, between two aligned sequences, in particular using algorithms as defined by Altschul *et al.*, *Nucleic Acids Res.* (1997) **25**:3389, for example.

The present invention relates in particular to a protein as defined above, in which the GFAT sequence corresponds to a bacterial or eukaryotic, in particular plant, fungal or animal, in particular insect or mammal, more particularly human GFAT sequence.

The invention relates more particularly to a protein as defined above, in which the purification tag sequence is inserted between two consecutive amino acids of the GFAT sequence, said amino acids being included in:

- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending between the β2 sheet and the β3 sheet of the *Escherichia coli* GFAT, or
- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending between the β13 sheet and the β14 sheet of the *Escherichia coli* GFAT, or
- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending between the β15 sheet and the α6 helix of the *Escherichia coli* GFAT.

The structure of the *Escherichia coli* GFAT is described in particular by Teplyakov *et al.*, *J. Mol. Biol.* (2001) **313**:1093 (whole protein), by Isupov *et al.*,

Structure (1996) **4**:801 (glutaminase domain) and by Teplyakov *et al.*, *Structure* (1998) **6**:1047 (isomerase domain). The structure of the complete protein can in particular be consulted using the 1JXA atomic coordinates file filed with the *Protein Data Bank* (http://www.pdb.org).

The *E. coli* GFAT peptide sequence is defined by SEQ ID NO: 13.

The sequence extending between the β2 sheet and the β3 sheet corresponds to the sequence extending approximately between amino acids 30 to 80 of *E. coli* GFAT, situated in the glutaminase domain.

The sequence extending between the β13 sheet and the β14 sheet corresponds to the sequence extending approximately between amino acids 220 to 230 of *E. coli* GFAT, situated in the glutaminase domain.

The sequence extending between the β15 sheet and the α6 helix corresponds to the sequence extending approximately between amino acids 235 to 250 of *E. coli* GFAT, situated in the isomerase domain.

According to a particular embodiment, the invention therefore relates to a protein as defined above, in which the purification tag sequence is inserted between two consecutive amino acids of the GFAT sequence, said amino acids being included in:

- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending approximately between amino acids 30 to 80 of *Escherichia coli* GFAT, or
- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending approximately between amino acids 220 to 230 of *Escherichia coli* GFAT, or
- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending approximately between amino acids 235 to 250 of *Escherichia coli* GFAT.

Identification of the parts of GFAT sequences corresponding and/or being homologous to secondary structures of *E. coli* GFAT can be obtained by aligning the sequence of said GFAT with that of *E. coli* GFAT, in particular using an algorithm such as that defined by Altschul *et al.*, *Nucleic Acids Res.* (1997) **25**:3389 or using the Clustal W software, well known to a person skilled in the art and described by Thompson *et al.*, Nucleic Acids Res. (1994) **22**: 4673-4680, for example.

In particular, two sequences or parts of sequences are referred to as homologous if the percentage of identity between these two sequences or parts of sequences is greater

than approximately 35% and/or if the percentage of similarity between these two sequences or parts of sequences is greater than approximately 44%.

More particularly, two sequences or parts of sequences are referred to as homologous if they are capable of hybridizing under stringent conditions, such as the following conditions: formamide 50%, NaCl 0.75 mol/l, sodium citrate 0.75 mmol/l, sodium dodecyl sulphate 1%, pH 7, 42°C.

According to another preferred embodiment, the invention relates to a protein as defined above, in which the purification tag sequence is inserted between two consecutive amino acids of a human GFAT sequence, said amino acids being included between amino acids 40 to 50, 290 to 330, and/or 340 to 370 of said human GFAT sequence.

Amino acids 40 to 50 of said human GFAT sequence correspond and/or are homologous to the part of the *E. coli* GFAT sequence extending between the β2 sheet and the β3 sheet, i.e. to the sequence extending approximately between amino acids 30 to 80 of *E. coli* GFAT.

Amino acids 290 to 330 of said human GFAT sequence correspond and/or are homologous to the part of the *E. coli* GFAT sequence extending between the β13 sheet and the β14 sheet, i.e. to the sequence extending approximately between amino acids 220 to 230 of *E. coli* GFAT.

Amino acids 340 to 370 of said human GFAT sequence correspond and/or are homologous to the part of the *E. coli* GFAT sequence extending between the β15 sheet and the α6 helix, i.e. to the sequence extending approximately between amino acids 235 to 250 of *E. coli* GFAT.

The invention relates in particular to a protein as defined above, in which the GFAT sequence corresponds to:

- SEQ ID NO: 2, corresponding to the human GFAT1 sequence,
- SEQ ID NO: 4, corresponding to the human GFAT2 sequence,
- SEQ ID NO: 6, corresponding to the human GFAT1Alt sequence.

The human GFAT1 sequence is in particular described in McKnight *et al.*, *J. Biol. Chem.* (1992) **267**:25208, and corresponds to the nucleotide sequence SEQ ID NO: 1.

The human GFAT2 sequence is in particular described in Oki *et al.*, *Genomics* (1999) **57**:227, and corresponds to the nucleotide sequence SEQ ID NO: 3.

The human GFAT1Alt sequence is in particular described in DeHaven *et al.*, *Diabetes* (2001) **50**:2419, and corresponds to the nucleotide sequence SEQ ID NO: 5.

The invention relates in particular to a protein as defined above, in which the purification tag sequence is inserted between two consecutive amino acids, said amino acids being included between amino acids:

- 43 to 47, 298 to 306, and/or 342 to 347 of SEQ ID NO: 2,
- 42 to 45, 299 to 307, and/or 343 to 348 of SEQ ID NO: 4
- 43 to 47, 316 to 324, and/or 360 to 365 of SEQ ID NO: 6

Amino acids 43 to 47 of SEQ ID NO: 2, 42 to 45 of SEQ ID NO: 4 and 43 to 47 of SEQ ID NO: 6 correspond, i.e. are homologous, to the part of the *E. coli* GFAT sequence extending between the β2 sheet and the β3 sheet, i.e. to the sequence extending approximately between amino acids 30 to 80 of *E. coli* GFAT.

Amino acids 298 to 306 of SEQ ID NO: 2, 299 to 307 of SEQ ID NO: 4 and 325 to 330 of SEQ ID NO: 6 correspond, i.e. are homologous, to the part of the *E. coli* GFAT sequence extending between the β13 sheet and the β14 sheet, i.e. to the sequence extending approximately between amino acids 220 to 230 of *E. coli* GFAT.

Amino acids 342 to 347 of SEQ ID NO: 2, 343 to 348 of SEQ ID NO: 4 and 360 to 365 of SEQ ID NO: 6 correspond, i.e. are homologous, to the part of the *E. coli* GFAT sequence extending between the β15 sheet and the α6 helix, i.e. to the sequence extending approximately between amino acids 235 to 250 of *E. coli* GFAT.

According to another particular embodiment, the invention relates more particularly to a protein as defined above, in which the purification tag sequence is inserted between two consecutive amino acids of the GFAT sequence, said amino acids being included in:

- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending approximately between amino acids 43 to 47 of human GFAT1,
- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending approximately between amino acids 298 to 306, in particular 299 to 300, of human GFAT1,
- a part of the GFAT sequence corresponding and/or being homologous to the sequence extending approximately between amino acids 342 to 347 of the human GFAT1.

The invention relates more particularly to an above protein, in which the purification tag sequence is inserted between amino acids:

- 299 and 300 of SEQ ID NO: 2,
- 300 and 301 of SEQ ID NO: 4,

- 317 and 318 of SEQ ID NO: 6.

The invention relates in particular to an above protein, in which the purification tag corresponds to a sequence of approximately 2 to approximately 10 amino acids, in particular of approximately 4 to approximately 8 amino acids.

Preferred purification tags according to the invention relate in particular to so-called FLAG tags (Sigma-Aldrich, France). These tags bind specifically to a given paratope, said paratope being able to belong to an antibody or to an antibody fragment for example. A particular example of a FLAG tag is constituted by the peptide sequence Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Asp-Lys (SEQ ID NO: 18) for example.

Other preferred tags according to the invention are tags formed by several histidines. These tags can form complexes with divalent metallic cations such as Ni^{2+} or Co^{2+} for example.

The invention relates in particular to a protein as defined above, in which the purification tag is a hexa-histidine.

The sequence His-His-His-His-His-His (SEQ ID NO: 19) is designated hexa-histidine.

The invention relates more particularly to a protein as defined above corresponding to the sequences:

- SEQ ID NO: 8, corresponding to the sequence SEQ ID NO: 2 in which a hexa-histidine is inserted between amino acids 299 and 300,
- SEQ ID NO: 10, corresponding to the sequence SEQ ID NO: 4 in which a hexa-histidine is inserted between amino acids 300 and 301, and
- SEQ ID NO: 12, corresponding to the sequence SEQ ID NO: 6 in which a hexa-histidine is inserted between amino acids 317 and 318.

The present invention also relates to a nucleic acid comprising or being constituted by a sequence coding for a protein as defined above.

The invention relates more particularly to a nucleic acid comprising or being constituted by the nucleotide sequence:

- SEQ ID NO: 7 coding for the protein SEQ ID NO: 8, or
- SEQ ID NO: 9 coding for the protein SEQ ID NO: 10, or
- SEQ ID NO: 11 coding for the protein SEQ ID NO: 12,

or by its complementary sequence, or being derived from said sequence by mutation, insertion or deletion of at least one nucleotide, provided that said nucleotide sequence codes for an enzymatically-active protein.

According to another embodiment, the present invention also relates to a eukaryotic or prokaryotic vector comprising a nucleic acid as defined above.

These vectors make it possible in particular to synthesize the proteins according to the invention in a eukaryotic or prokaryotic organism.

Advantageously the invention relates to a baculovirus-type expression vector allowing the synthesis of the proteins according to the invention in insect cells.

The present invention also relates to a process for the purification of a protein as defined above, from a solution comprising said protein, comprising a stage of bringing said solution into the presence of a compound binding specifically to the purification tag of said protein and a stage of separation of the complex formed by the binding of said protein to said compound from the other constituents of the solution.

The compound can be fixed to a solid support such that the complex formed between said compound and said protein can be recovered by centrifugation or filtration. Optionally said compound fixed on its support can be arranged in a column through which said solution is eluted.

Advantageously, the above process can also comprise a stage of dissociation of the complex formed by the binding of said protein to said compound in order to recover the purified protein.

The invention relates more particularly to a purification process as defined above, comprising a stage of bringing a solution comprising a protein as defined above into the presence of a compound comprising a divalent metallic cation such as Ni^{2+} or Co^{2+}, in particular Ni^{2+}, and a stage of separation of the complex formed by the binding of the protein to said compound from the other constituents of the solution.

Advantageously, the above process can also comprise a stage of dissociation of the complex formed by the binding of said protein to said compound comprising a divalent metallic cation, in particular using imidazole, in order to recover the purified protein.

According to another embodiment the present invention relates to a process for the preservation of a protein as defined above in an enzymatically-active form, in particular at -80°C or at 4°C, comprising the addition of said protein to a solution comprising:

- approximately 1 mM to approximately 10 mM of fructose 6-phosphate, in particular approximately 1 mM,

- approximately 1 mM to approximately 5 mM of Tris(2-carboxyethyl)phosphine, in particular approximately 1 mM,
- approximately 5% to approximately 20% of glycerol, in particular approximately 10%.

The fructose-6-phosphate is a substrate of said protein.

Tris(2-carboxyethyl)phosphine is a reducing compound advantageously making it possible to maintain the property of resins carrying Ni^{2+} or Co^{2+} ions.

Advantageously the glycerol is a cryoprotective agent.

The present invention therefore also relates to a composition comprising an active GFAT protein, if appropriate, bound to a purification tag, such as a protein as defined above, said protein being capable of being preserved in an enzymatically-active form, for at least 8 days at a temperature of 2°C to 10°C, in particular approximately 4°C, and for at least 12 months at a temperature of -100°C to -20°C, in particular approximately -80°C, said protein being in combination with:

- approximately 1 mM to approximately 10 mM of fructose 6-phosphate, in particular approximately 1 mM,
- approximately 1 mM to approximately 5 mM of Tris(2-carboxyethyl)phosphine, in particular approximately 1 mM,
- approximately 5% to approximately 20% of glycerol, in particular approximately 10%.

The present invention also relates to the use of a protein as defined above, for the screening of compounds modifying the activity of said protein, in particular for the screening of said protein inhibitor.

The activity of the proteins according to the invention can in particular be measured using the following methods:

- the radiometric method described by Broschat *et al.*, *Analytical Biochem.* (2002) **305**:10-15,
- the so-called Nitro Blue Tetrazolium method described by Nakata *et al.*, *J. Antibio.* (2001) **54**:737-743.
- the Morgan-Elson method described by Ghosh *et al.*, *Method. Enzymol.* (1960) **5**:414 and described in detail in the example which follows.
- the APAD method described by Badet *et al.*, *Biochemistry* (1987) **26**:1940 and described in detail in the example which follows.

Advantageously these methods can be used for the screening, in particular at a high flow rate, of compounds modifying the activity of the proteins according to the invention.

The invention relates in particular to use as defined above, for the screening of compounds useful within the framework of the treatment or prevention of diabetes, in particular type II diabetes, obesity, acidosis, ketosis, arthritis, cancer, or osteoporosis.

DESCRIPTION OF FIGURE 1

Figure 1 represents the plasmid pFastBac-gfat-His6 with a molecular weight of 6.89 kb. The cassette "Ampr" represents an ampicillin-resistance gene, the cassette "ori" represents a bacterial replication origin, the cassette "Gmr" represents a gentamicin-resistance gene, the cassette "Polh Pr" represents the polyhedrin promoter, the cassette gfat-his6 represents the gfat1 gene modified by the insertion of a sequence coding for a hexahistidine tag. The *Xba*I restriction sites in position 4.11 kb, and *Eco*RI in positions 4.56 kb and 6.60 kb are also represented.

EXAMPLE 1

1. Synthesis and cloning of the *gfat1-His6* gene

The *Eco*RI fragment of a cDNA corresponding to the human *gfat1* gene was cloned in the *Eco*R I site of the pCRII vector (Invitrogen) in order to form the plasmid pCRII-gfat1. The nucleotide sequence of an internal purification tag composed of 6 histidine residues was introduced at position 898 of the *gfat1* gene sequence cloned in pCRII by PCR with *Platinum pfx polymerase* (Roche) and the appropriate pair of primers:

- Start Aat II-His6:

5'TGGACGTCTTTCTATCCATCGAATTAAACGAACTGCAGGACATCACC ATCACCATCACGATCACCCCGGACG 3' (SEQ ID NO: 14)

- End Hinc II:

5' CAAAGTTGACTCTTCCTCTCATTGTGTTCACGACAGACTCTGGC 3' (SEQ ID NO: 15)

according to the following protocol: 94°C, 2 minutes then 30 cycles (94°C 45 seconds, 55°C 1 minute, 72°C 5 minutes) followed by 5 minutes' polymerization at 72°C and returning to 4°C.

After digestion by *Aat*II and *Hinc*II then purification on 1.5% Seaplaque agarose gel (Tebu), the amplicon (170 bp) was inserted at the level of the corresponding restriction sites into the pCRII-gfat1 construction. The 170 bp insert was introduced by ligation into the construction with a ratio of 3:1 at 16°C overnight in the presence of T4 DNA ligase (Nebs). The ligation mixture (20 µl) thus obtained made it possible to transform a strain of *E. coli* JM109. Then, the *Xba*I- *Hind*III fragment of the recombinant plasmid pCRII-gfat1-His6 was cloned in the donor plasmid pFastBac1 (Life Technologies Ltd). The plasmid pFastBac-gfat-His6 thus generated (**Figure 1**) was verified by multiple digestions: *Sma*I, *Acc*I/*Dra*I, *Pst*EI/*Xba*I, and by sequencing. With a view to improving the construction, the sequence upstream of the start codon was mutated at two positions by PCR, with the following pair of primers, in order to remove two open reading frames upstream of the *gfat1* gene:

- Start XbaI

5' AATCTAGATTCATGCTCGAGCGGCCGCCAGTGTGATTGATATC 3' (SEQ ID NO: 16)

- End AfeI

5' ATTTTTATCAGAGCGCTGGGGGTGGCTATTGACAGG 3' (SEQ ID NO: 17)

according to the protocol: 94°C 2 minutes, then 30 cycles (94°C 15 seconds, 55°C 30 seconds, 68°C 1 minute) followed by a 1 minute's polymerization at 68°C and returning to 4°C.

The PCR fragment obtained, containing the two mutations, was purified on SeaPlaque gel (Tebu) at 0.7% then digested by *Xba*I and *Afe*I in order to replace its homologue in pFastBac-gfat-His6 in order to produce the donor plasmid pFastBac-gfat-His6-2orf to be used for transposition into the DH10Bac cells (Life Technologies Ltd). The construction was verified by *Sma*I, *XbaI*/*Pst*EI, *Xba*I/*Hind*III digestions, and by sequencing.

A recombinant bacmid was isolated after transposition into the DH10Bac cells and used for transfecting Sf9 insect cells in the presence of Lipofectin (Life Technologies Ltd). The baculoviruses obtained were amplified in the Sf9 cells and the viral titre was measured at 5.10^7 pfu/ml.

2. Production of the GFAT1-His6 protein

Sf9 insect cells were cultured at 28°C in the presence of SF900II medium (Life Technologies Ltd) in 5 l flasks under stirring at 100 rpm. The cells at a density of 2.10^9 cells/l were infected by the recombinant baculovirus obtained above with an infection multiplicity of 2 (pfu/cell), then cultured for 72 hours.

The cells and the supernatant were separated by centrifugation (2500 g, 10 minutes at 4°C). The cell pellets were washed in the presence of 20 mM Tris-HCl buffer, pH 7, centrifuged (4000 g, 45 minutes at 4°C) and frozen at –80°C.

3. Purification of the GFAT1-His6 protein

The cell pellet (20 g) was taken up in 50 ml of lysis buffer (50mM $NaPO_4$, pH 7.5, 300 mM NaCl, 10 mM imidazole, 1 mM fructose-6-phosphate (fructose-6P), 1 mM TCEP (Tris(2-carboxyethyl)phosphine), 1 mM PMSF (phenylmethylsulphonyl fluoride), 10% glycerol and 1 protease inhibitor cocktail tablet without EDTA (Roche Applied Sciences) and subjected to grinding with a DynoMill at 4500 rpm (4 cycles of 30 seconds) in the presence of 40 g of microbeads 0.2 mm in diameter. The mixture was cooled down by circulation of ethylene glycol/water adjusted to -15°C. The crude

extract obtained (100 ml, 445 mg of total proteins) was centrifuged at 4°C for 20 minutes at 12000 rpm. The supernatant was subjected to ultracentrifugation at 4°C (350,000 rpm, 1 hour). The supernatant thus obtained was mixed with 5 ml of 50% Ni-NTA matrix (Qiagen) for 2 hours at 4°C. The mixture was poured into an empty column then rinsed with 40 ml of washing buffer (50 mM $NaPO_4$, pH 7.5, 300 mM NaCl, 40 mM imidazole, 1 mM fructose-6P, 1 mM TCEP, 1 mM PMSF and one protease inhibitor cocktail tablet without EDTA (Roche Applied Sciences). Elution was carried out by successive stages with 125 and 500 mM imidazole in the same buffer as previously. 12 mg of functional GFAT1-His6 (protein assay according to Bradford's method) were thus obtained.

4. Preservation of the GFAT1-His6 enzyme

The enzyme was then stored in 100 µl fractions in the presence of 1 mM fructose-6P, 1mM TCEP and 10% glycerol at –80°C. The stability of the enzyme is several months at –80°C and more than 8 days at 4°C.

5. Assay of the GFAT1-His6 enzyme activity

Different assay tests of the enzyme activity of GFAT1-His6 were used. These tests can be also used in order to screen compounds modifying, and in particular inhibiting, the activity of the GFAT1-His6. It is possible to easily adapt them to screening at a high flow rate.

Morgan-Elson Assay:

In this case the enzyme activity is monitored by a colorimetric test the principle of which is the following: the D-glucosamine-6P released by the enzyme is N-acetylated by acetic anhydride in alkaline medium (Ghosh *et al.*, *Method. Enzymol.* (1962) **5**:414), then the solution is treated with Ehrlich's reagent (para-dimethyl-amino-benzaldehyde, PDAB) in concentrated acid medium; the pink compound formed absorbs at 585 nm.

The enzymatic reaction takes place over 30 minutes at 37°C in the presence of:

- 0.2 ml of 100 mM fructose-6P
- 0.25 ml of 60 mM L-Glutamine
- 0.25 ml of 150 mM KPO_4 buffer, pH 7
- 0.1 ml of 25 mM EDTA (ethylene diamine tetra-acetate), pH 7

- up to 200 µl of sample (to be completed with H_2O if necessary)

The reaction is stopped by immersion for 4 minutes in a water bath at 100°C then centrifuged. 0.8 ml of the supernatant is sampled for assay of the glucosamine-6P according to the following protocol:

- addition of 0.1 ml of saturated $NaHCO_3$,
- addition of 0.1 ml of a 5% acetic anhydride solution in water prepared extemporaneously,
- stirring and incubation for 5 minutes at ambient temperature,
-incubation for 5 minutes in a bath at 100°C,
- addition of 0.2 ml of 0.8 M potassium borate, pH 9.1 (to be adjusted with 10 N KOH).
-stirring and incubation for 7 minutes in a bath at 100°C.
-addition of 3 ml of Ehrlich's reagent diluted 10 times in acetic acid, prepared extemporaneously, to the solution cooled down in ice,
- incubation for 20 minutes at 37°C.

The activity of the GFAT was determined by comparison with a standard curve established using D-glucosamine as standard in a concentration range of 0 to 200 nmoles. The specific activity of the GFAT1-His6 obtained was thus measured at 1.7 U/mg. This is greater than the value of 0.4 U/mg obtained by Broschat *et al.*, *J. Biol. Chem.* (2002) **277**:14764, for the purification of a recombinant human GFAT1. This reflects a greater activity of the GFAT1-His6 and/or a greater purity of the enzyme preparation according to the invention.

The kinetic parameters of GFAT1-His6 have been characterized vis-à-vis glutamine (K_m^{Gln} = 0.2 mM) and fructose-6P (F6P) (K_m^{F6P} = 0.006 mM) by a spectrophotometric assay coupled with glutamate dehydrogenase according to the APAD test. This is in accordance with the values cited in the prior art (K_m^{Gln} = 0.26 mM and K_m^{F6P} = 0.007 mM for Broschat *et al.*, *J. Biol. Chem.* (2002) **277**:14764).

APAD assay

This is an ultraviolet spectrophotometric assay of the GFAT activity. It is based on the continuous determination of the quantity of L-glutamate formed using GFAT and an analogue of NAD (nicotinamide adenine dinucleotide), APAD (3-acetylpyridine

adenine dinucleotide), according to the following reaction (catalyzed by glutamate dehydrogenase (GDH)):

$$\text{L-Glu} + \text{APAD} \xrightarrow[\text{GDH}]{} \alpha\text{- ketoglutarate} + NH_4^+ + \text{APADH}$$

The measurement is carried out at 365 nm, at 37°C. Under these conditions an absorbance unit corresponds to 0.11 µmole of APADH formed.

The test comprises:

- 100 µl 3 mM APAD (2 mg/ml)
- 25 µl 2M KCl
- 100 µl of 1 M KPO_4 buffer, pH 7.2
- 100 µl of 100 mM Fructose-6P (30.41 mg/ml)
- 100 µl of 60 mM purified L-Glutamine (8.77 mg/ml)
- H_2O qsf 1 ml (taking into account the volumes to be added hereafter)
- 50 µl GDH
- sample to be assayed: 0.5 µg

It is also possible to use other assay processes, such as the radiometric assay described by Broschat *et al.*, *Analytical Biochem.* (2002) **305**:10-15 or the so-called Nitro Blue Tetrazolium assay described by Nakata *et al.*, *J. Antibio.* (2001) **54**:737-743.